

Meteorological DB and XML DB

Toshiyuki Amagasa
Computational Intelligence Group
Computational Informatics Division

amagasa@cs.tsukuba.ac.jp



The GPV/JMA Archive


<http://gpvjma.ccs.hpcc.jp>

GPV/JMA Archive: Home - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り 移動

アドレス(D) <http://gpvjma.ccs.hpcc.jp/~gpvjma/>



GPV/JMA Archive

Data by Japan Meteorological Agency
Contents Provided by the Center for Computational Sciences
University of Tsukuba

[HOME](#)
[REGISTER](#)
[ARCHIVE](#)
[e-mail](#)

About the archive

This Archive offers the daily operational weather forecasting data provided by the Japan Meteorological Agency (JMA). The data are called Grid Point Values (GPV). The Archive is maintained by the Center for Computational Sciences, University of Tsukuba, for the purpose of scientific development of the weather and climate forecasting technology. All weather maps posted here are the product by the CCS, University of Tsukuba, Japan

Files stored

In the Archive, there are six kinds of JMA/GPV data, i.e., global spectral model data (gsm_jma), regional spectral model data (rsm_jma), meso-scale non-hydrostatic model data (msm_jma), weekly ensemble forecast data (ensemble_week_jma), monthly ensemble forecast data (ensemble_month_jma), and seasonal ensemble forecast data (ensemble_3month_jma). Those GPV data are stored in subdirectories describing the date (yyyymmdd00) when the data are generated. The dated subdirectories are combined in the main directory describing the year.

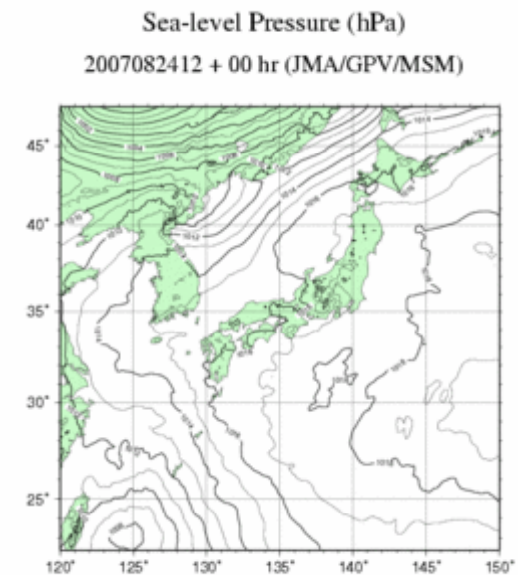
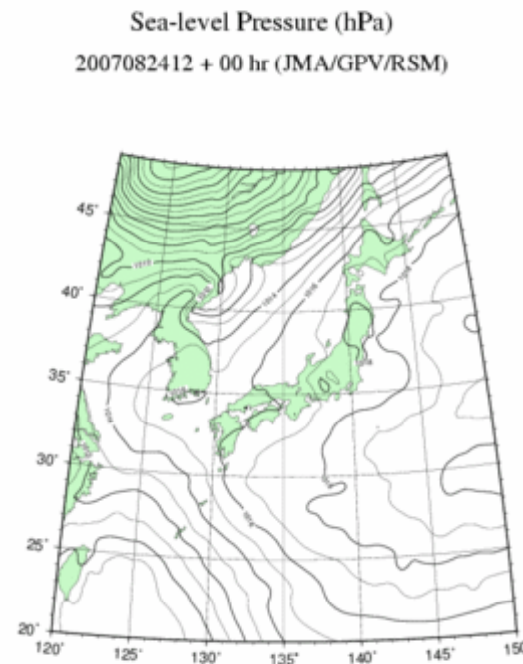
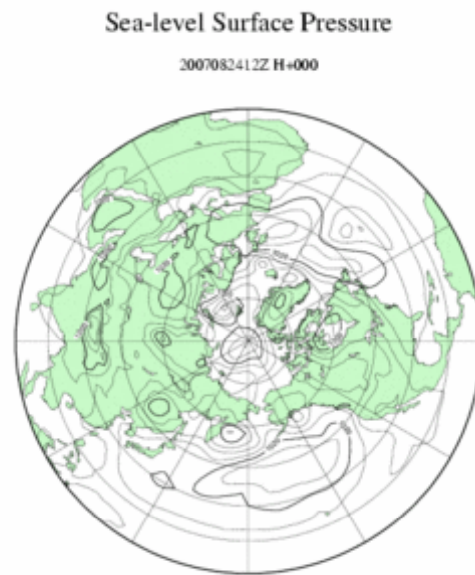
Notice: Due to the system upgrade of JMA, the resolution and format of the GPV data have changed after March 1, 2006. Refer to the appropriate documents issued by JMA.

ページが表示されました インターネット



GPV data by JMA

- ▶ Global Spectral Model (GSM)
- ▶ Regional Spectral Model (RSM)
- ▶ Meso-Scale non-hydrostatic Model (MSM)

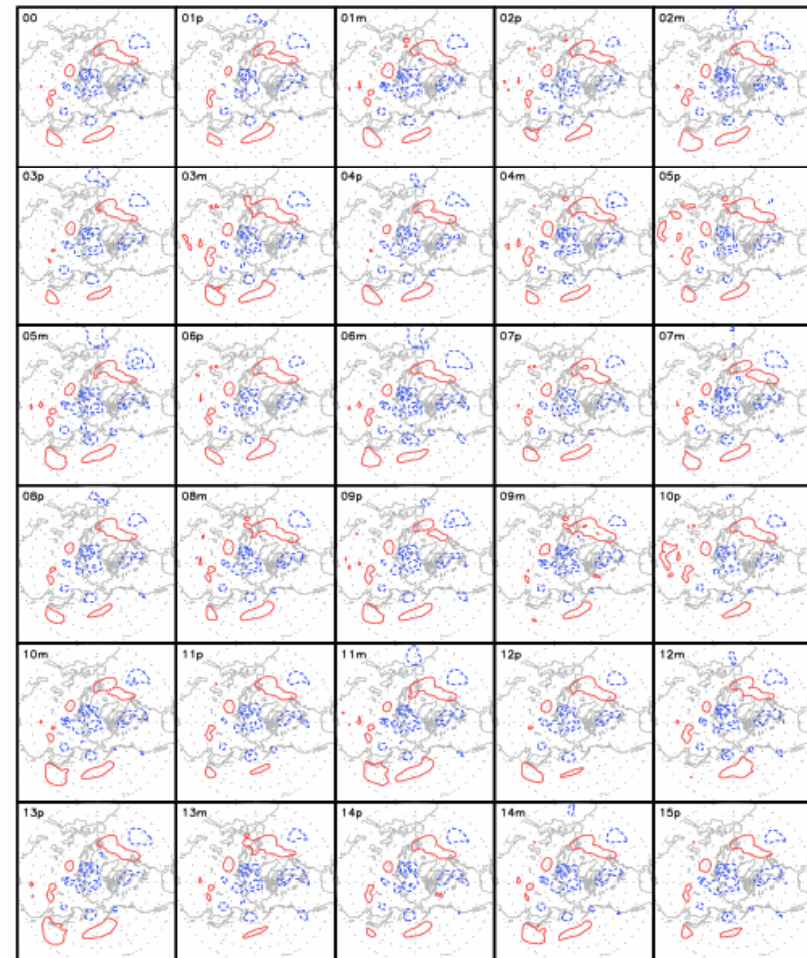




GPV data by JMA (cont'd)

- ▶ Weekly ensemble
- ▶ Monthly ensemble
- ▶ Seasonal ensemble

JMA Week Ensemble Forecast (PRMSL)
Anomaly 20070824 12UTC +000hr





Meteorological DBs

Japan

- Kitsueregawa, Toyoda lab., The University of Tokyo
 - <http://www.tkl.iis.u-tokyo.ac.jp:8080/GPV/>
- Center for Climate System Research, The University of Tokyo
 - <http://www.ccsr.u-tokyo.ac.jp/>
- Disaster Prevention Research Institute, Kyoto University
 - http://www.dpri.kyoto-u.ac.jp/web_j/

Other countries

- National Centers for Environmental Prediction
 - <http://www.cdc.noaa.gov/cdc/reanalysis/>
- European Centre for Medium-Range Weather Forecasts
 - <http://data.ecmwf.int/data/>
- National Oceanic and Atmospheric Administration
 - <http://www.ngdc.noaa.gov/stp/SOLAR/solar.html>
- CISL Research Data Archive
 - <http://dss.ucar.edu/>



Background

▶ e-Science

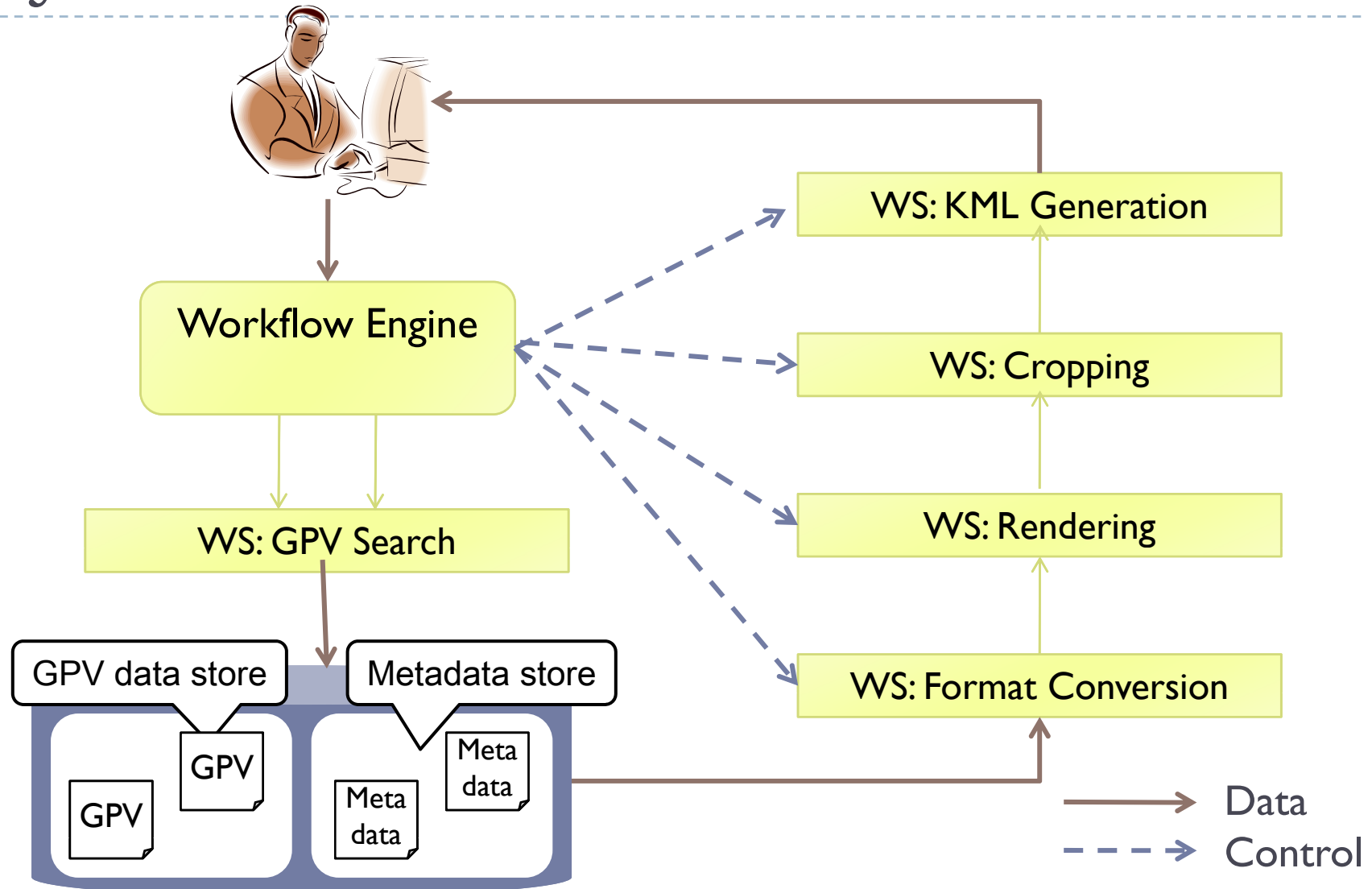
- ▶ Applying ICT technologies for accelerating
 - ▶ Exchanging scientific resources
 - ▶ Sharing computational power
 - ▶ Human interaction

▶ Objective

- ▶ Constructing e-Science infrastructure for meteorology
 - ▶ Web service system for supporting domain experts
 - DB query
 - Data acquisition
 - Format conversion
 - Rendering for visual analysis
 - ...



System overview



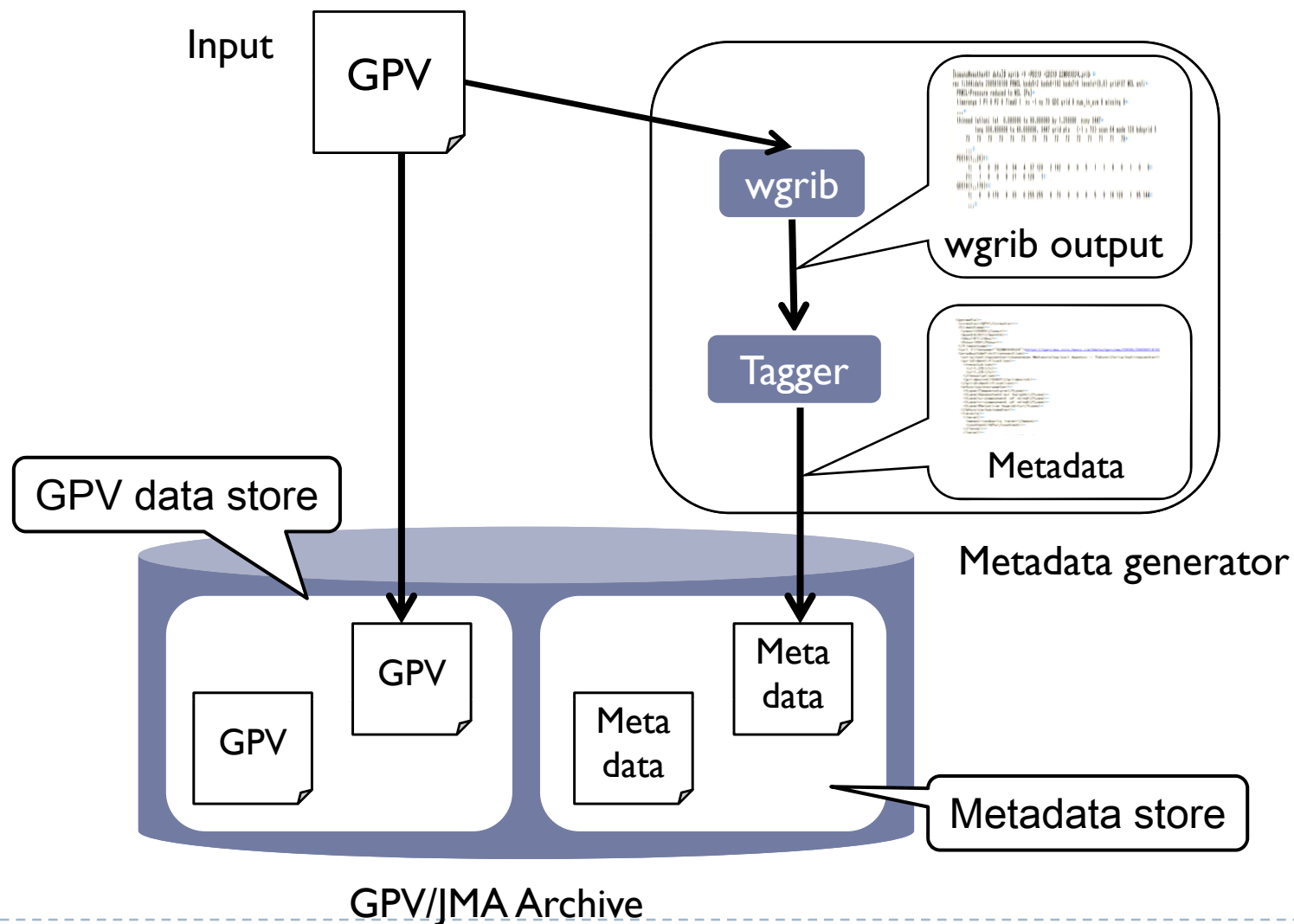


Metadata design for GPV/JMA data

- ▶ **Using XML**
 - ▶ Compatibility with other related formats
 - ▶ Query processing by XPath / XQuery
- ▶ **Design policy**
 - ▶ Inherent information
 - ▶ Generated timestamp
 - ▶ Data generator
 - ▶ Types of physical attributes (pressure, temperature, wind vector, ...)
 - ▶ Grid information (resolution, number, width, ...)
 - ▶ Surface information (constant-pressure, ...)
 - ▶ Projective method
 - ▶ Additional information
 - ▶ URL
 - ▶ Data format (GSM, RSM, MSM)



Metadata extraction





Metadata extraction (cont'd)

- ▶ Generate XML data from dumped GPV data

```
[komano@weather01 data]# wgrib -V -PDS10 -GDS10 GSM00X024.grib ←
rec 1:344: date 2005010100 PRMSL kpds5=2 kpds6=102 kpds7=0 levels=(0,0) grid=37 MSL anl: ←
PRMSL=Pressure reduced to MSL [Pa] ←
timerange 1 P1 0 P2 0 Timell 1 nx -1 ny 73 GDS grid 0 num in ave 0 missing 0 ←
... ←
thinned lat ... 0.0 ... lon 330.0 ... 60.000000, ... grid pts ... x 73 scan 64 mode 128 bdsgrid 1
73 73 73 ... 73 73 73 ... 72 72 71 71 71 70 ←
... ←
PDS10(1..28)= ←
1: 0 0 28 3 34 4 37 128 2 102 0 0 5 1 1 0 0 1 0 0 ←
21: 1 0 0 0 0 21 0 128 1 ←
GDS10(1..178)= ←
1: 0 0 178 0 33 0 255 255 0 73 0 0 0 0 5 9 16 128 1 95 144 ←
... ←
```

TimeStamp

Creator

GridIdentification

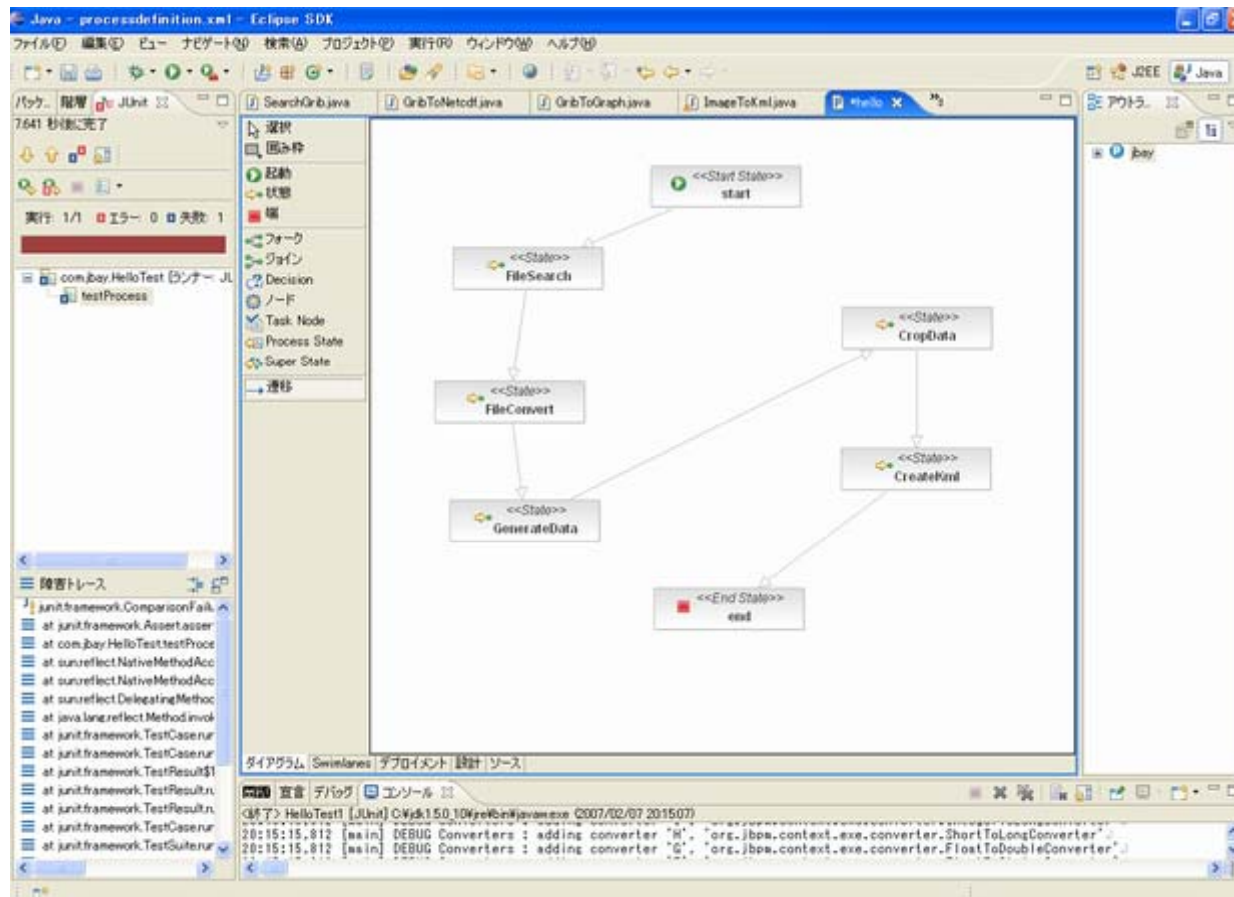
PhysicalParameter

Level

RepresentationType

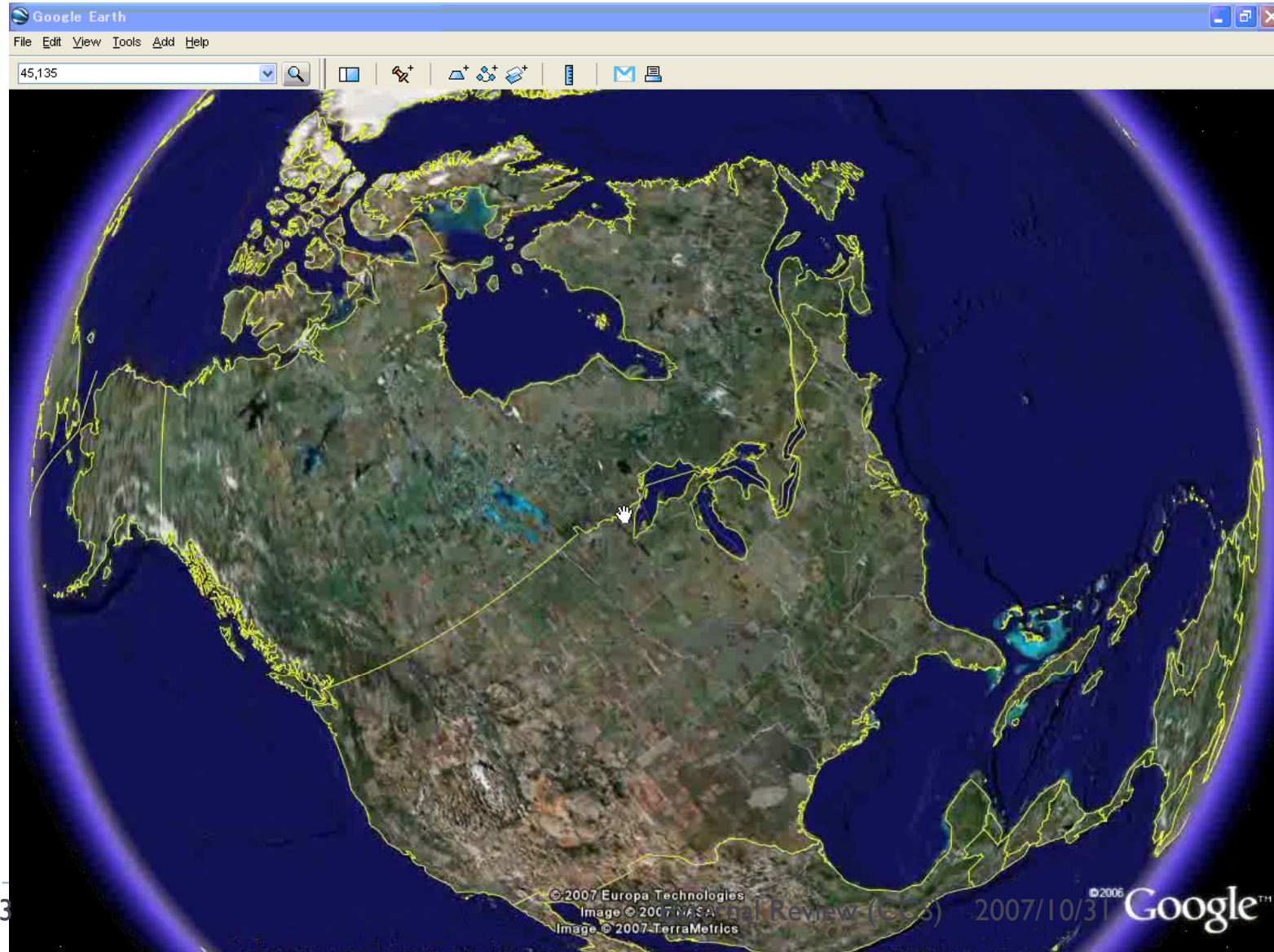


Workflow specification by BPEL4WS





Demo





Funding and output

▶ Funding

- ▶ Grant-in-Aid for Exploratory Research by JSPS

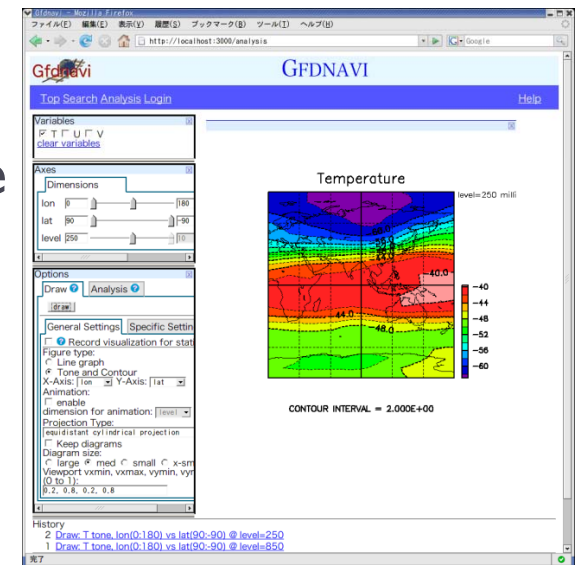
▶ Publication

- ▶ T.Amagasa, H. Kitagawa, and T. Komano, "Constructing a Web Service System for Large-scale Meteorological Grid Data", 3rd IEEE Int'l Conf. on e-Science and Grid Computing (e-Science 2007), Bangalore, India, Dec. 10-13, 2007.



Collaborations

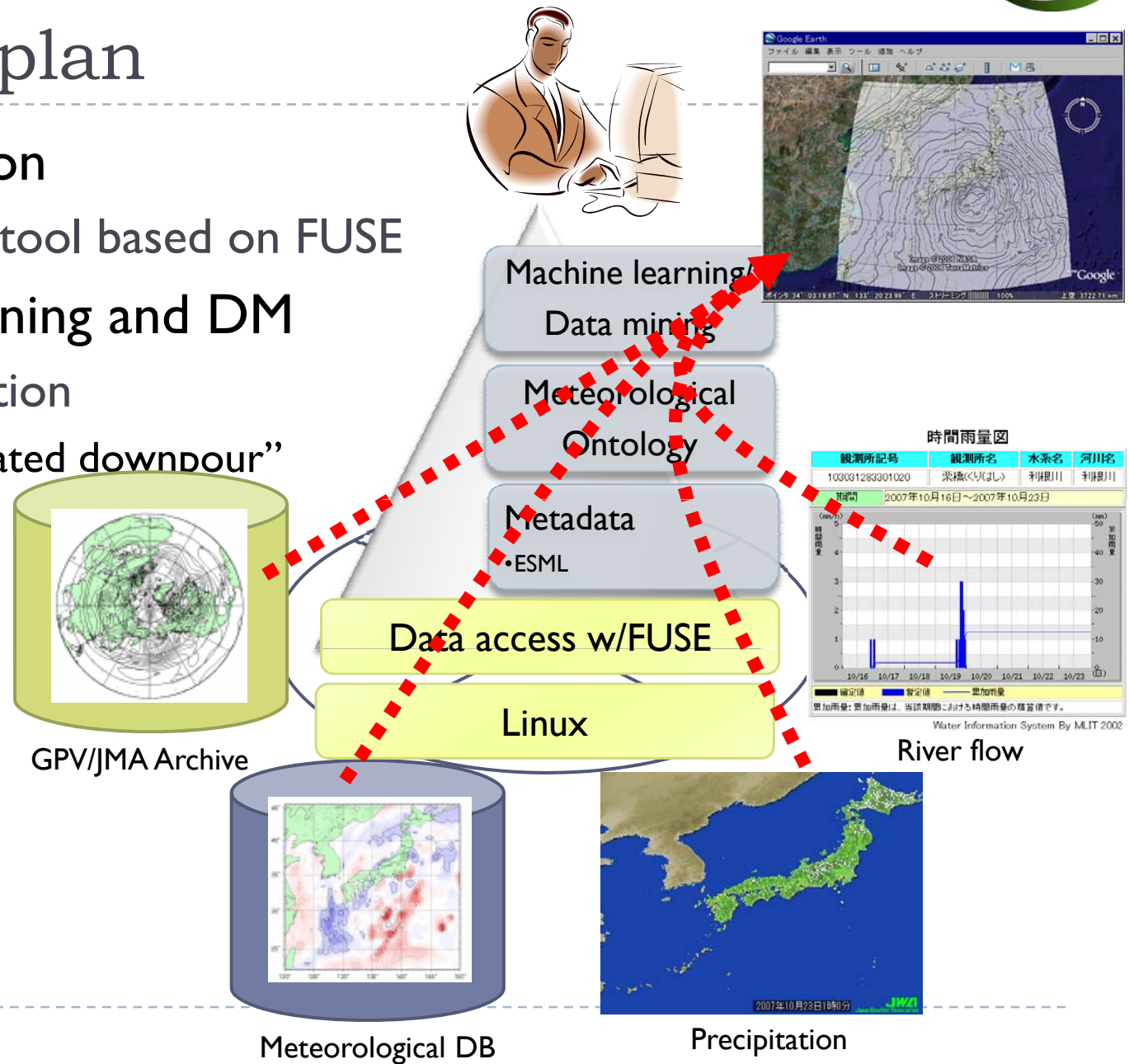
- ▶ Inside CCS
 - ▶ Global Environmental Science Group
- ▶ External
 - ▶ GeoGrid (AIST)
 - ▶ Massive sensing data from satellites (ASTER & MODIS)
 - ▶ gfdnavi
 - ▶ Toolkit for earth scientists
 - ▶ National Institute for Land and Infrastructure Management (NILIM)
 - ▶ Sensing data of river flow
 - ▶ Comprehensive tie-up agreement between Univ. of Tsukuba and NILIM





Research plan

- ▶ DB Integration
 - ▶ Lightweight tool based on FUSE
- ▶ Machine learning and DM
 - ▶ Event detection
 - ▶ “Concentrated downpour”



XML Databases



XML: Standard data format on wire

- ▶ **Meta language for data representation**
 - ▶ Represent tree structure by plain text
- ▶ **Rapidly diffusing**
 - ▶ AML, BML, CML, ..., ZML, AAML, ABML, ACML, ...
- ▶ **Typical usecases**
 - ▶ Web documents
 - ▶ XHTML
 - ▶ Business documents
 - ▶ ODF
 - ▶ MS Office Open XML
 - ▶ Metadata
 - ▶ RDF (blogs)
 - ▶ Scientific data

```
<CATALOG>
  <CD>
    <TITLE>Empire Burlesque</TITLE>
    <ARTIST>Bob Dylan</ARTIST>
    <COUNTRY>USA</COUNTRY>
    <COMPANY>Columbia</COMPANY>
    <PRICE>10.90</PRICE>
    <YEAR>1985</YEAR>
  </CD>
  ...
</CATALOG>
```



Research objective and recent activities

► Providing infrastructure for managing massive XML resources

Integration of heterogeneous XML resources

- Proximity queries over heterogeneous XML data
- Similarity join

Analytical processing of XML data

- OLAP for XML

Storage and retrieval of XML in various information systems

- PC cluster systems
- P2P networks



Analytical processing of XML

How to extract information from XML?

▶ Searching

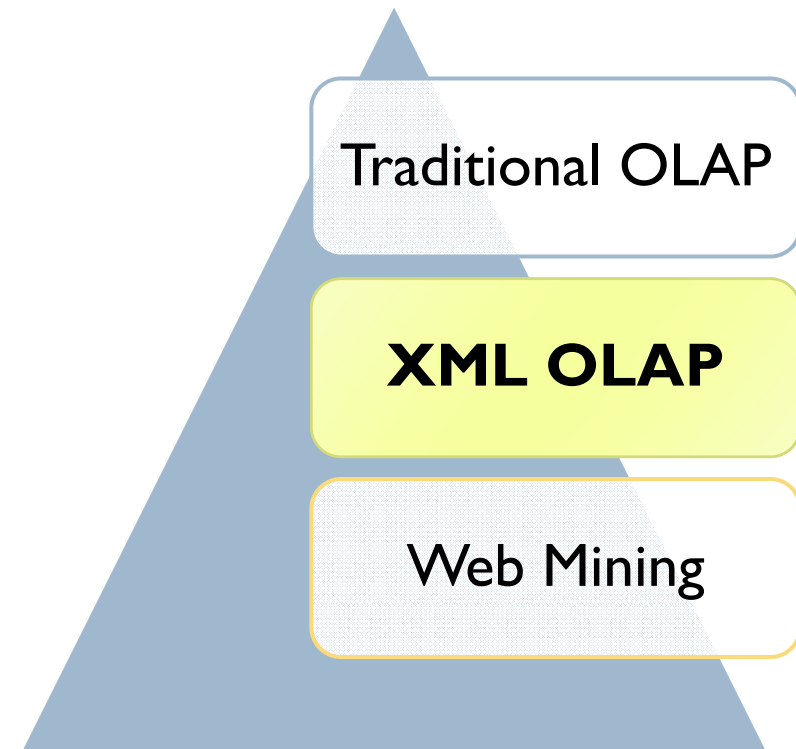
- ▶ Path-based query languages
 - ▶ XPath and XQuery

▶ Information retrieval

- ▶ Keyword-based queries
 - ▶ CO (Content Only)
 - ▶ CAS (Content and Structure)

▶ Analytical processing

- ▶ OLAP-like interactions
 - ▶ Drilling, Cubing, Pivoting, ...



➔ Interactive analysis for mining useful information



Example: DBLP Bibliography

```
<inproceedings mdate="2002-01-03" key="conf/er/AmagasaATK98">  
<author>Toshiyuki Amagasa</author>
```

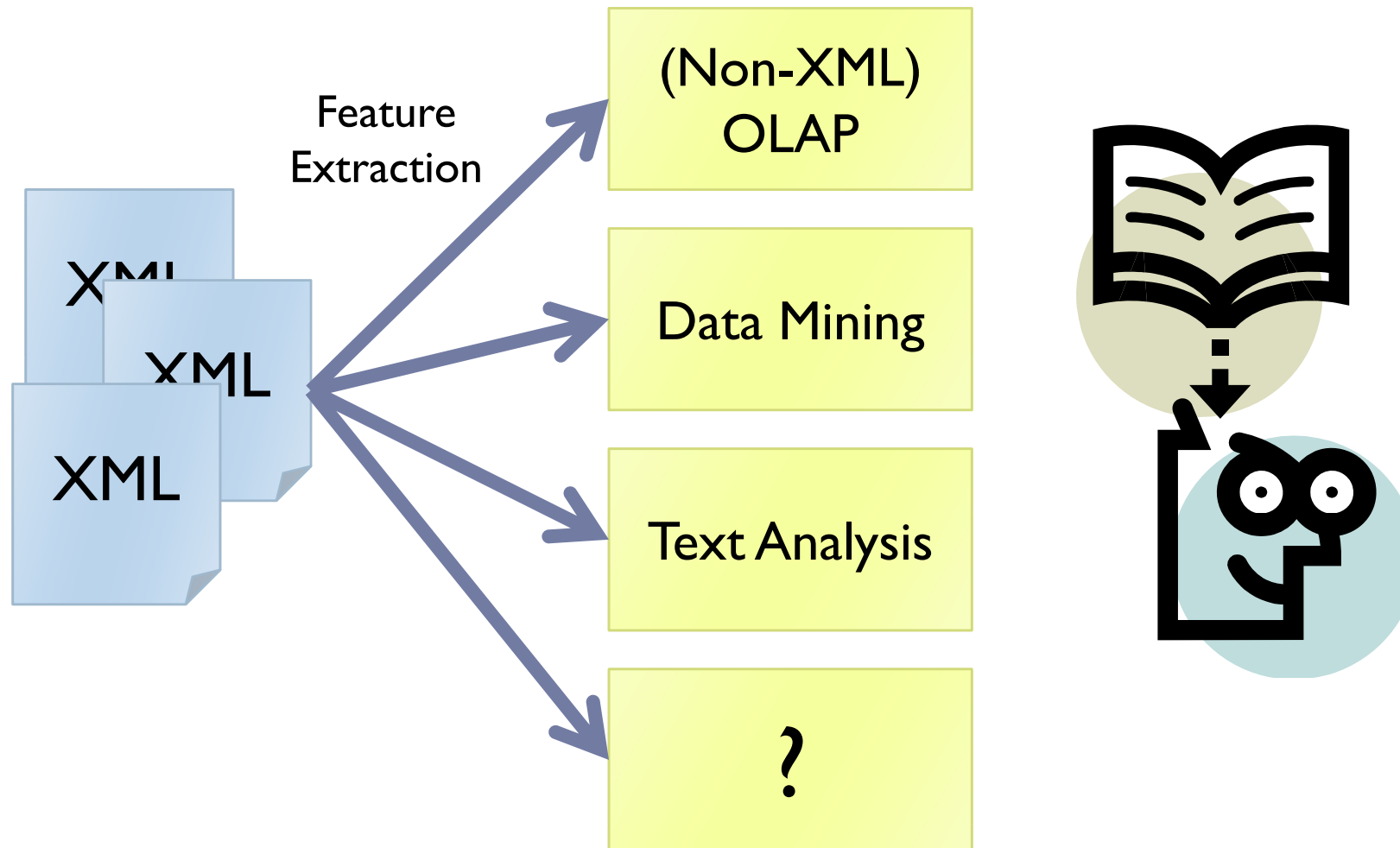
Typical examples of analysis: “Research trends in the DB domain”

- Number of submitted papers
 - Topic
 - Journal / Conference
- Number of distinct authors
 - Topic
 - Journal / Conference
- Number of frequently used keywords
 - Topic
 - Journal / Conference
 - Author
- ...

```
</inproceedings>  
<cite>conf/dexa/AritsugiTAK97</cite>
```

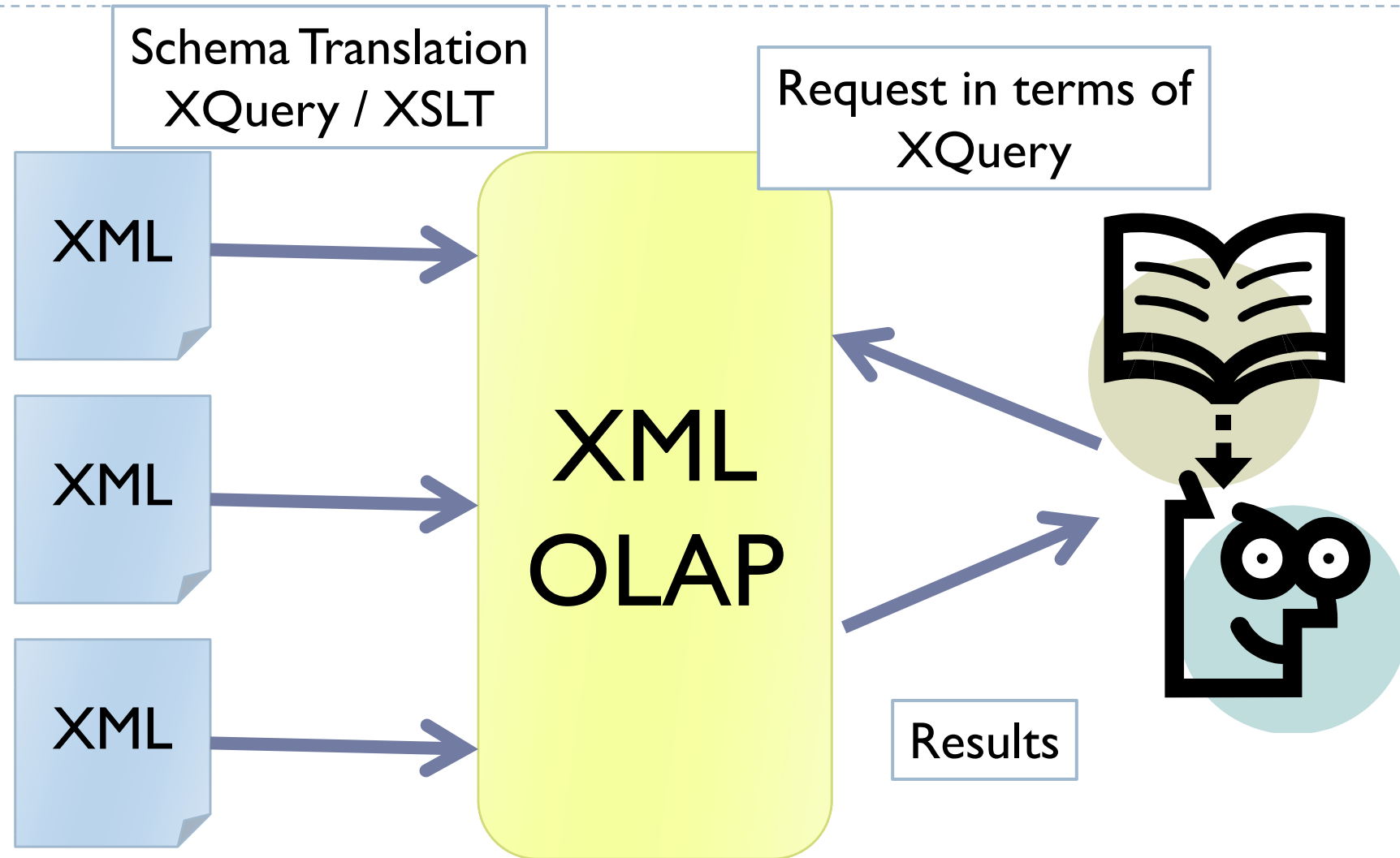


Traditional way





XML-OLAP way

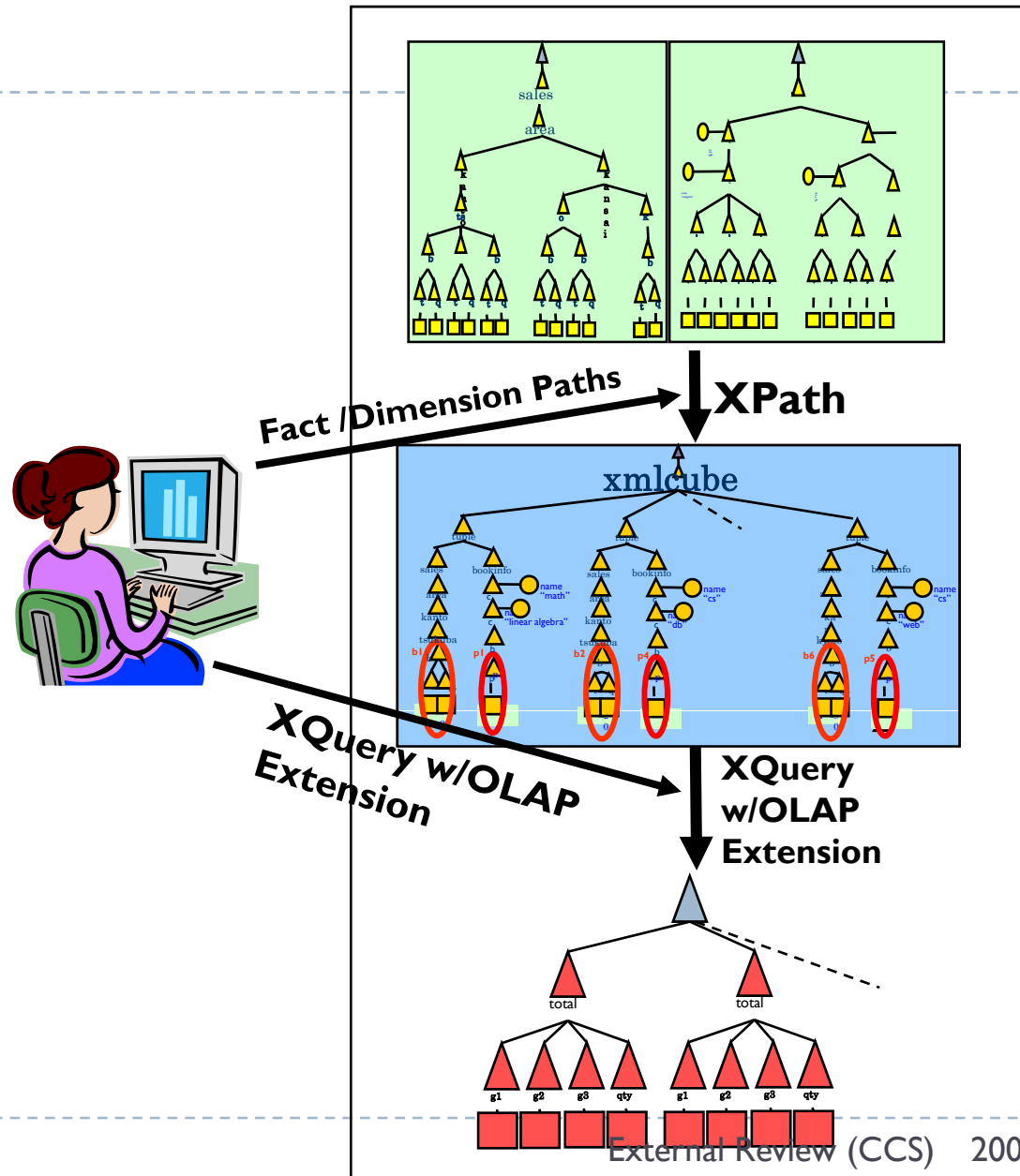




Technical challenges

- ▶ **Supporting concept hierarchy**
 - ▶ XML itself contains hierarchies for XML-OLAP
- ▶ **Supporting non-numerical attributes**
 - ▶ Only numerical attributes are supported in (Non-XML)OLAP
 - ▶ XML contains both numerical and **text** attributes
 - ▶ How we can summarize text according to concept hierarchy?
- ▶ **Heterogeneity of XML data**
 - ▶ Irregular structure
 - ▶ Missing attributes / values
 - ▶ Heterogeneous schemas

Proposed system



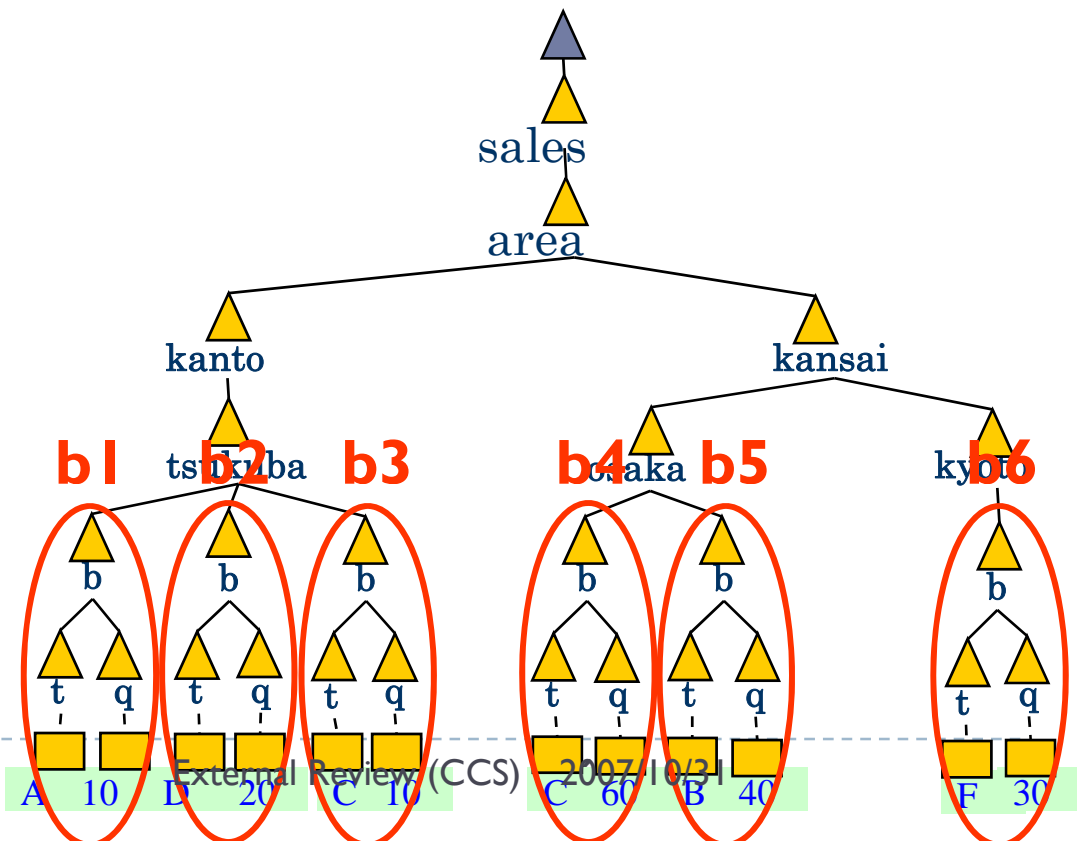


Facts about an XML Data

- ▶ Facts in an XML data are data items of interest, identified by an XPath (fact path).

Ex. A user wants to get information of book sale by giving:

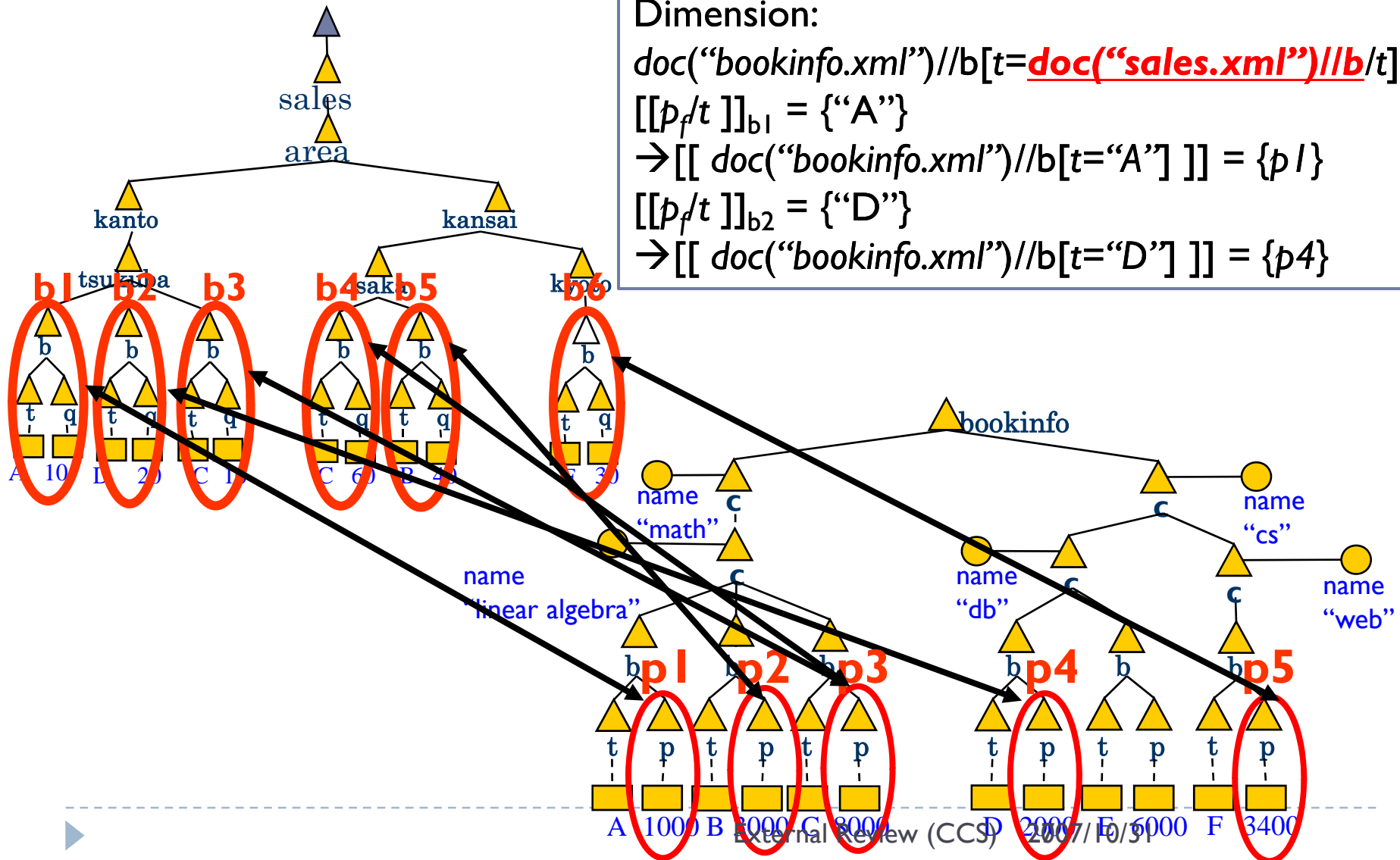
p_f : doc("sales.xml")//b
Facts: $[[p_f]] = \{b1, \dots, b6\}$



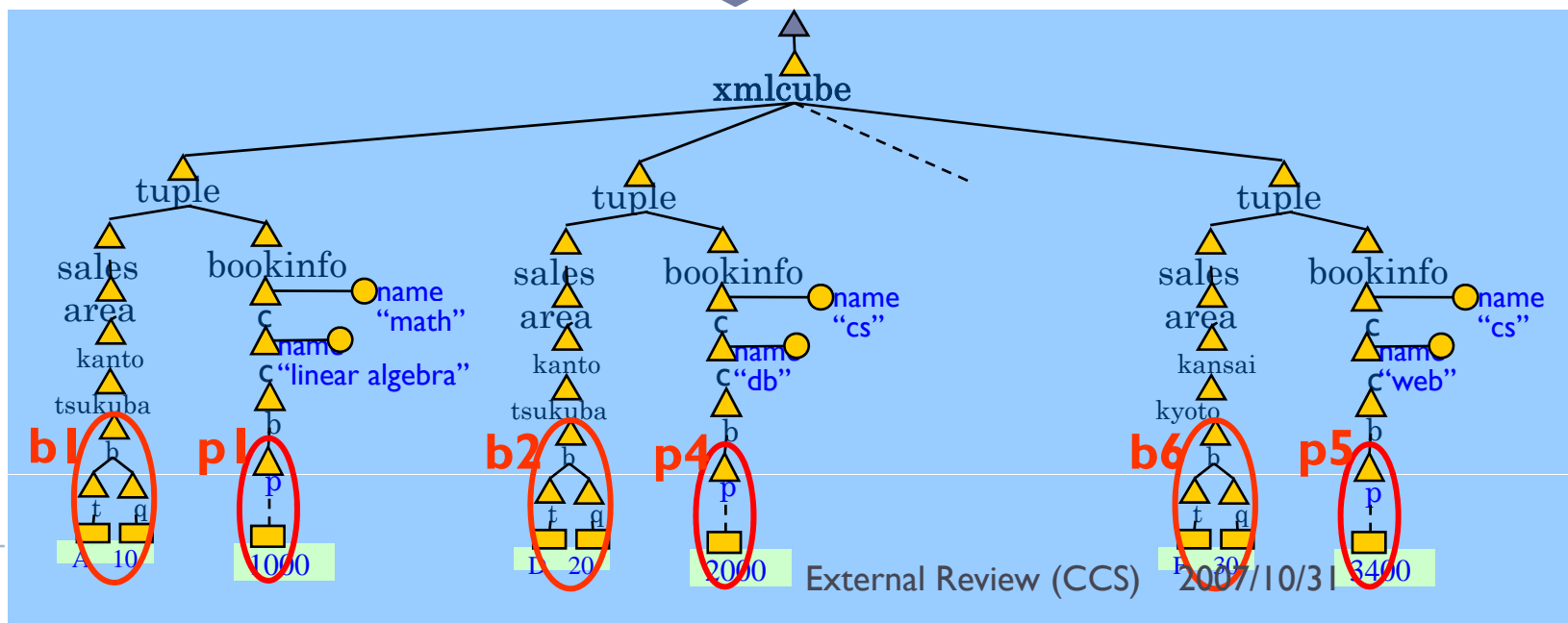
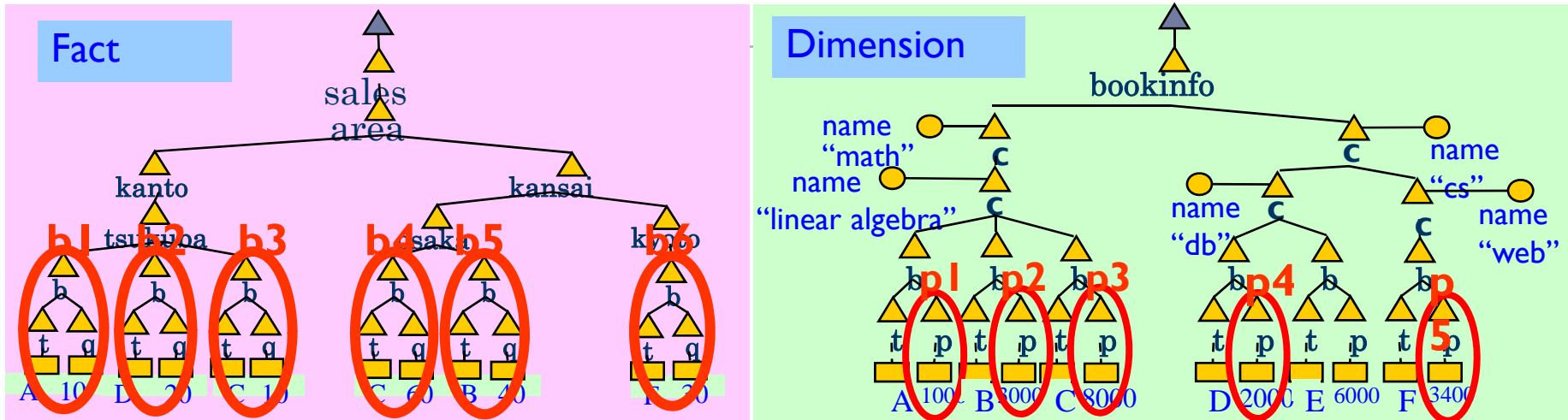


Dimensions

- ▶ A dimension is given as a path expression (dimension path)

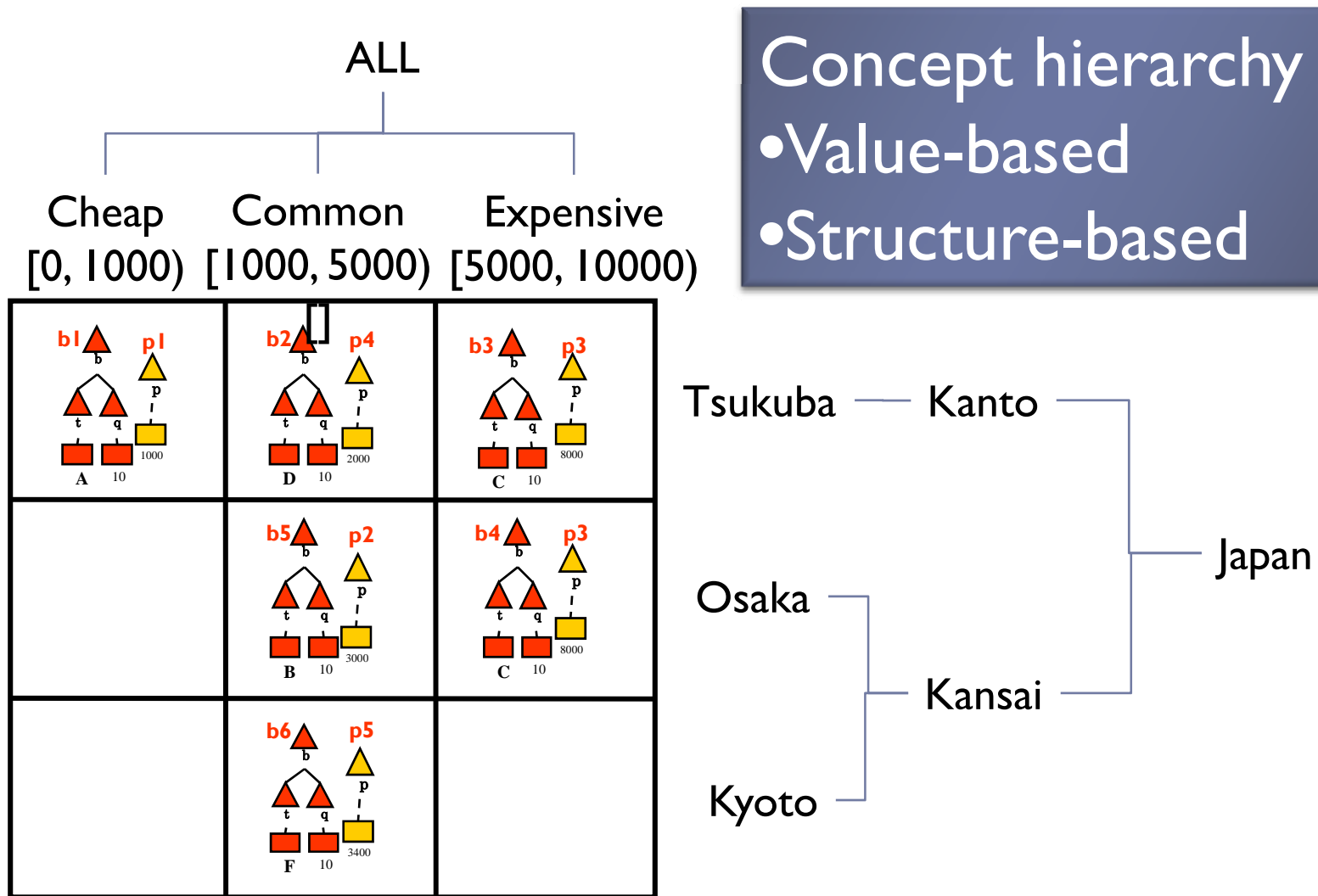


XML Cube

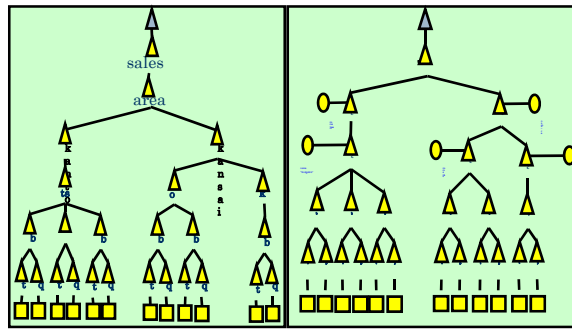




XML-Cube and concept hierarchy



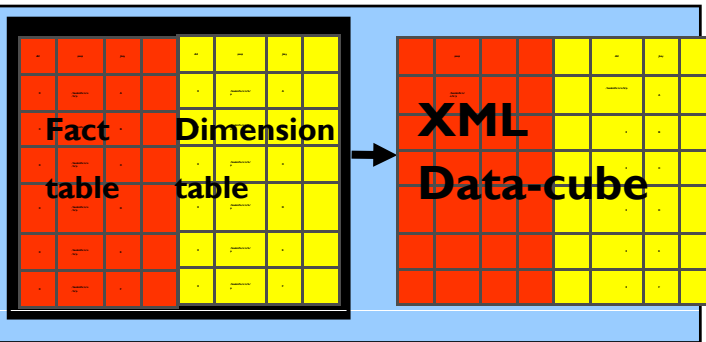
System



Implementation

Path Approach

id	pid	cid	name	value
1	1	1		
2	1	2		
3	1	3		
4	1	4		
5	1	5		
6	1	6		
7	1	7		
8	1	8		
9	2	9		
10	2	10		
11	2	11		
12	2	12		
13	3	13		
14	3	14		
15	3	15		
16	3	16		
17	3	17		
18	3	18		
19	3	19		
20	3	20		
21	3	21		
22	3	22		
23	3	23		
24	3	24		
25	3	25		
26	3	26		
27	3	27		
28	3	28		
29	3	29		
30	3	30		
31	3	31		
32	3	32		
33	3	33		
34	3	34		
35	3	35		
36	3	36		
37	3	37		
38	3	38		
39	3	39		
40	3	40		
41	3	41		
42	3	42		
43	3	43		
44	3	44		
45	3	45		
46	3	46		
47	3	47		
48	3	48		
49	3	49		
50	3	50		
51	3	51		
52	3	52		
53	3	53		
54	3	54		
55	3	55		
56	3	56		
57	3	57		
58	3	58		
59	3	59		
60	3	60		
61	3	61		
62	3	62		
63	3	63		
64	3	64		
65	3	65		
66	3	66		
67	3	67		
68	3	68		
69	3	69		
70	3	70		
71	3	71		
72	3	72		
73	3	73		
74	3	74		
75	3	75		
76	3	76		
77	3	77		
78	3	78		
79	3	79		
80	3	80		
81	3	81		
82	3	82		
83	3	83		
84	3	84		
85	3	85		
86	3	86		
87	3	87		
88	3	88		
89	3	89		
90	3	90		
91	3	91		
92	3	92		
93	3	93		
94	3	94		
95	3	95		
96	3	96		
97	3	97		
98	3	98		
99	3	99		
100	3	100		



Fact /Dimension Paths

XPath

Query Translation

SQL

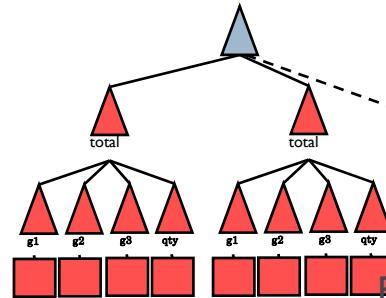


XQuery w/OLAP Extension

XQuery w/OLAP Extension

Query Translation

SQL

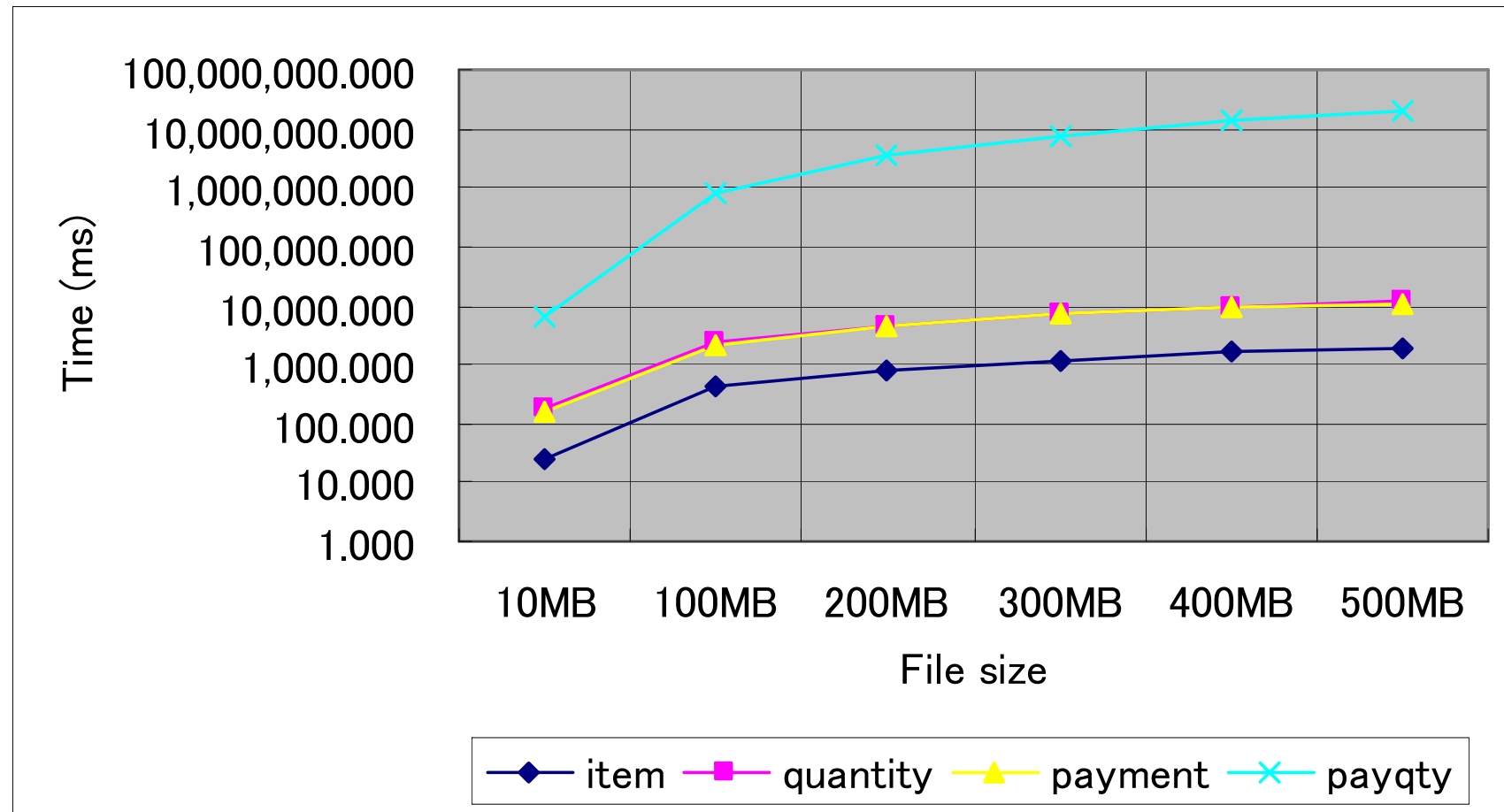


g1	g2	g3	Total
1
2	CreditCard_Cash	Cash	164
2	CreditCard_Cash	CreditCard	164
2	CreditCard_Cash	CreditCard	164
2	CreditCard_Personal Check	CreditCard	133
2	CreditCard_Personal Check	Personal Check	133
2	CreditCard_Personal Check	Personal Check	133
2	Money order_Cash	Cash	143
2	Money order_Cash	Money order	143
2	Money order_Cash	Money order	143
...

External Review (CCS)

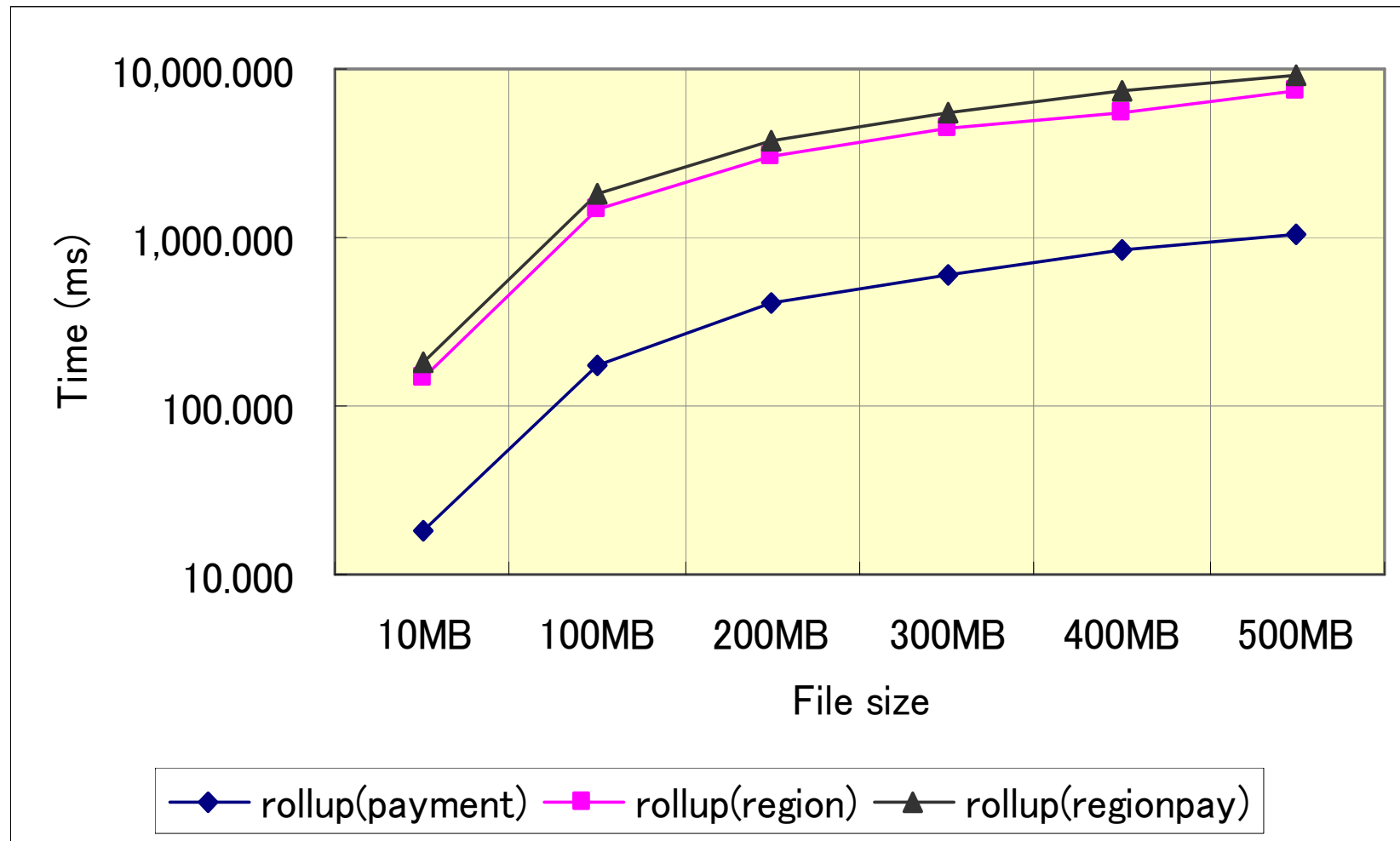


Time for XML-cube construction





Time for GROUP BY computation





Related work

- ▶ **XML-OLAP is becoming a hot research area.**
 - ▶ R. Bordawakar and C.A. Lang. Analytical Processing of XML Documents: Opportunities and Challenges. In SIGMOD Record, volume 34(2), pages 27–32, 2005.
 - ▶ Chaitanya Gokhale, Nitin Gupta, Pranav Kumar, Laks Lakshmanan, Raymond Ng, B. Aditya Prakash: Complex Group-by Queries for XML, ICDE2007, 2007.
 - ▶ Nuwee Wiwatwattana, H.V. Jagadish, Laks Lakshmanan, Divesh Srivastava: X^3 : A Cube Operator for XML OLAP, ICDE2007, 2007.



Conclusions

- ▶ **Proposed XML-OLAP**
 - ▶ Facts and dimensions of XML data
 - ▶ Prototype implementation using relational storage
 - ▶ Experimental evaluation
- ▶ **Future work**
 - ▶ Algorithms for computing multiple GROUP BYs by taking XML features into account
 - ▶ System implementation and in-depth analysis





Award and outputs

▶ Award

- ▶ DBSJ Paper Award 2005: "Processing Large-Scale XML Data by RDBMS using Region Directory"

▶ Recent publications

- ▶ Toshiyuki Amagasa, Lianzi Wen, and Hiroyuki Kitagawa, "Proximity Search of XML Data using Ontology and XPath Edit Similarity", 18th Int'l Conference on Database and Expert Systems Applications (DEXA2007), pp. 298-307, Regensburg, Germany, Sept., 2007.
- ▶ Toshiyuki Amagasa, Kentarou Kido, and Hiroyuki Kitagawa, "Querying XML Data using PC Cluster System", 2nd International Workshop on XML Data Management Tools and Techniques (XANTEC'07) in conjunction with DEXA 2007, pp. 5-9, Regensburg, Germany, 2007.
- ▶ Toshiyuki Amagasa, Chunhui Wu, and Hiroyuki Kitagawa, "Retrieving Arbitrary XML Fragments from Structured Peer-to-Peer Networks," Joint Conference of the 9th Asia-Pacific Web Conference and the 8th International Conference on Web-Age Information Management (APWeb/WAIM 2007), pp. 317-328, HuangShan? (Yellow Mountains), China, June 2007.
- ▶ Chantola Kit, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "OLAP Query Processing for XML Data in RDBMS," The Third IEEE International Workshop on Databases for Next-Generation Researchers (SWOD2007) in conjunction with ICDE2007, Istanbul, Turkey, April 15, 2007.
- ▶ Kentarou Kido, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "Processing XPath Queries in PC Clusters Using XML Data Partitioning," The Second International Special Workshop on Databases for Next-Generation Researchers (SWOD 2006) in conjunction with ICDE 2006, pp. 114-119, Atlanta, Georgia, USA, April 7, 2006.
- ▶ Kei Fujimoto, Toshiyuki Shimizu, Dao Dinh Kha, Masatoshi Yoshikawa and Toshiyuki Amagasa, "A Mapping Scheme of XML Documents into Relational Databases using Schema-based Path Identifiers," International Workshop on Challenges in Web Information Retrieval and Integration (WIRI 2005), April 8-9, 2005